

Tipo de trabalho: Trabalho individual ou em grupos de até 4 pessoas.

Data limite de entrega: 22/11/2005

Objetivo: realizar uma aplicação de análise de regressão e apresentá-la na forma de um artigo.

Sobre os dados:

Podem ser levantados ou obtido de fontes secundárias.

- Devem haver pelo menos duas variáveis quantitativas e uma qualitativa.
- Devem haver no mínimo 40 observações.

Informações:

1) A análise de regressão consiste em obter uma equação matemática que relacione uma variável dependente quantitativa (ou transformada em quantitativa) com uma ou mais variáveis independentes, também quantitativas, (cujos valores poderiam causar um efeito na variável dependente). A obtenção da equação consiste em definir a *forma* da curva de regressão e os *parâmetros* desta curva. Por exemplo, pode-se utilizar uma *reta*, sendo necessário calcular os *coeficientes* da reta.

2) Quando há variáveis *qualitativas* na base de dados, e deseja-se incluí-las no modelo de regressão, utiliza-se o expediente de criar variáveis "dummy" associadas às variáveis qualitativas. As variáveis "dummy" são variáveis binárias, que podem assumir apenas valores 0 ou 1, e que são dessa forma incluídas no modelo. Maiores detalhes no manual eletrônico do Minitab.

3) Quando há apenas uma variável independente a regressão é dita simples. Quando há mais de uma ela é dita múltipla.

4) É possível realizar análise de regressão utilizando vários programas computacionais: Excel, Minitab, entre outros.

Requisitos do trabalho:

1) Introdução. Com a colocação do problema, objetivo da análise, descrição das variáveis (indicando qual é a dependente, e qual ou quais são independentes, se haverá variáveis "dummy", etc.). Por exemplo, queremos obter um modelo de regressão das vendas de refrigerante na temporada de verão em função das informações de temperatura e umidade relativa do ar.

2) Desenvolvimento. Aplicação da análise de regressão propriamente dita, que deverá incluir os seguintes aspectos (incluindo os gráficos e tabelas necessários, que podem ser construídos através do Statistica 5.1):

2.1 - Definição da equação de regressão. Por exemplo, uma equação linear:

$$\text{Vendas} = a \times \text{Temperatura} + b \times \text{Umidade} + c$$

2.2 - Apresentação de testes de hipóteses sobre os parâmetros. Por exemplo, testar, a partir dos dados, se os parâmetros a e b da equação mostrada no item 2.1 são diferentes de zero (ou seja se há regressão).

2.3 - Análise dos resíduos, diferença entre os valores *observados* da variável dependente e os valores *preditos* da mesma variável, através do modelo (frente às variáveis independentes e frente aos valores preditos pela equação de regressão): interpretação da adequação do modelo. Não pode haver padrões não aleatórios nos diagramas de dispersão (scatterplot) dos resíduos, e eles precisam apresentar uma distribuição normal (o que pode ser verificado pelo normal probability plot).

2.4 - Indicação das transformações eventualmente aplicadas aos dados (caso a análise de resíduos indique que as suposições para utilização do modelo foram violadas). Realizar novamente o item 2.3 para mostrar que os resíduos indicam que o modelo agora está adequado.

2.5- Equação final do modelo de regressão (incluindo, se necessário as transformações aplicadas no item 2.4).

2.6 - Interpretação dos resultados: como é a forma e a força do relacionamento entre as variáveis, e o que se pode esperar de previsões feitas a partir do modelo.

3) Considerações finais: pequeno resumo do trabalho, síntese das constatações obtidas pelo trabalho.

4) Bibliografia consultada.

5) Arquivo de dados. Pode ser disponibilizado em disquete, NÃO HÁ NECESSIDADE DE IMPRIMIR OS DADOS.

Condições para entrega:

- O trabalho deve ser entregue IMPRESSO e a equipe deverá apresentá-lo (máximo de 20 minutos) para a turma (a apresentação também fará parte da avaliação).

- O trabalho deve ser entregue e apresentado na data estipulada. Eventuais atrasos implicarão em reduções proporcionais na nota do trabalho.

- Os alunos devem procurar o professor da disciplina para dirimir eventuais dúvidas, além de consultar a bibliografia sugerida abaixo:

BARBETTA, P.A. Estatística Aplicada às Ciências Sociais. 5ed - Florianópolis: Ed. Da UFSC, 2004, Capítulo 13.

TRIOLA, M. Introdução à Estatística. Rio de Janeiro: LTC - Livros Técnicos e Científicos Editora S.A., 1999, Capítulo 9.

Minitab - Manual eletrônico.

LEVINE, D. M., BERENSON, M. L. e STEPHAN, D. – Estatística: Teoria e Aplicações usando o Excel. Rio de Janeiro: LTC, 2000, Capítulos 11 e 12.

Fontes de dados:

<http://lib.stat.cmu.edu/> - Biblioteca estatística da Universidade Carnegie Mellon.

<http://exploringdata.cqu.edu.au/datasets.htm> - Dados no site Exploring Data.

<http://www.mat.univie.ac.at/~neum/statdat.html> - Dados estatísticos no site da Universidade de Viena.

<http://www.sci.usq.edu.au/staff/dunn/Datasets/tech-regression.html> - Dados estatísticos no site da Universidade de Southern Queensland, Austrália.

Marcelo Menezes Reis

1/11/2005